Pictorial Depth Cues for Outdoor Augmented Reality

Jason Wither, Tobias Höllerer
University of California, Santa Barbara
{jwither, holl}@cs.ucsb.edu

Abstract

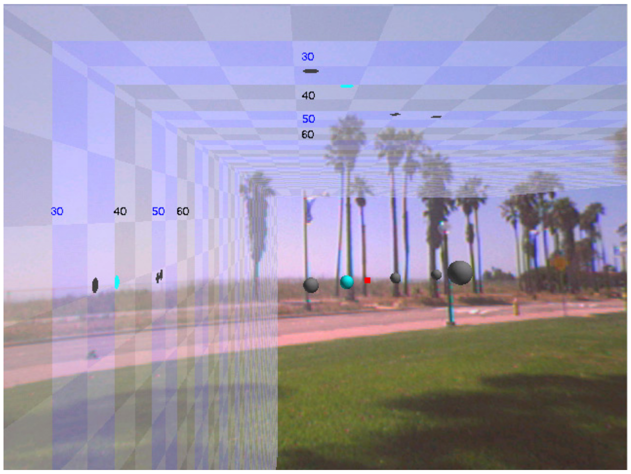
This paper presents and evaluates a set of pictorial depth cues for far-field outdoor mobile augmented reality (AR). We examine the problem of accurately placing virtual annotations at physical target points from a static point of view. While it is easy to line up annotations with a target point's projection in the view plane, finding the correct distance for the annotation is difficult if the target point is not represented in an environment model. We have found that AR depth cues, such as vertical and horizontal shadow planes, a small top-down map, or color encodings of relative depth, have a positive impact on a user's ability to align a 3D cursor with physical objects at various distances. These cues aid the user's depth perception and estimation by providing information about the 3D cursor's distance and its relationship in 3-space to any features that may already have been annotated. We conducted a user study that measures the effects of different depth cues for both absolute 3D cursor placement as well as placement relative to a small number of marked reference points, whose distances are known. Our study provides insight about mobile AR users' ability to judge distances both absolutely and relatively, and we identify techniques that successfully enhance their performance.

1. Introduction

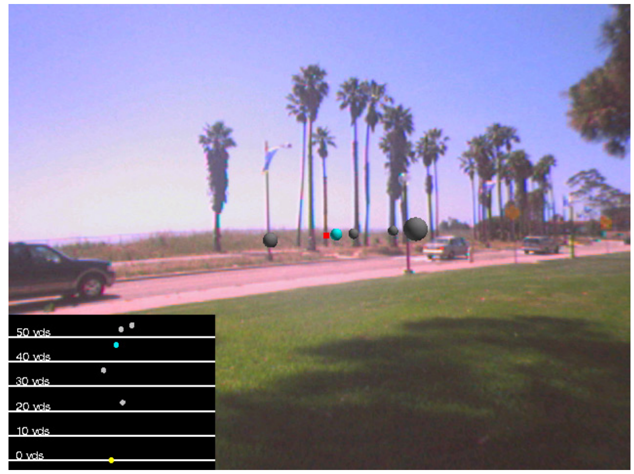
Mobile augmented reality (AR) is a powerful interface paradigm for wearable computing, both as an output medium for world-registered visualizations [9] as well as an input medium for annotating and modeling the physical world [2][15]. The ability to attach information to physical objects is an important concept in location-aware wearable computing [1][21]. This can be done by map- and model-based offline authoring or by the mobile user interacting with the physical world [19]. Assuming that the mobile user is position-tracked, information can be placed by walking to a certain spot and virtually dropping information there. If a scene model of the environment surrounding the user is available, virtual reality techniques for interaction at a distance can be used to place information on top of objects in the user's view. Here we examine annotation placement in the very common case where no scene model is available and the user is some distance away from the object to be annotated. As a motivating example, consider the scenario of [23], in which a mobile user annotates and classifies trees from a distance efficiently creating a geo-registered tree inventory.

For any given user location, it is easy to align annotations with the projections of physical features onto the view plane. However, in order to place an annotation in world coordinates, the correct distance between user and object is needed. In absence of a scene model, a mobile user could "sight down" the physical object with intersecting viewing rays or planes shot into the scene from different locations, requiring the user to strategically choose viewing angles and walk considerable distances for accurate results. This paper examines the question of how well a *static* user can estimate distances using AR aids to help them select the correct depth. The techniques and evaluations presented here are therefore appealing complements to the modeling techniques described in [2][15][16][17].

We have conducted a user study comparing different cues to aid the user in placing an annotation at the correct distance from the user. Our cues help the user judge the distance to objects in the real world, especially when there are already some objects in the user's view annotated in the model (which can be seen as the first few coordinates in an evolving scene model). We developed and tested four different AR cues for this purpose. The first required the user to judge distance given only the size of the 3D cursor and other markers in the scene. The second cue introduced two shadow planes where orthographic projections of virtual makers in the scene are cast. The third cue used a small top-down view to show the user the relationship between the cursor and other marked objects in the scene. The last cue showed how the cursor's depth is related to that of existing markers by having them change color based on their relative distance along the viewer's viewing direction. Users navigated a cursor through the 3D scene using these cues until they were satisfied their cursor was at the correct location for annotation. For this cursor movement we used a finger-mounted trackball and the most successful placement technique from our previous study on interaction at a distance [23]. In order to conduct an accurate study, we also needed to improve our registration beyond the results obtained by pure inertial/magnetometer based orientation tracking. To do



(a) Shadow Planes



(b) Top-Down View

(c) Color Encoded Markers

Figure 1: Three test cases from our study. The user controls the light blue cursor to annotate the tree marked by the small red dot. These images illustrate the second part of the study, evaluating depth control relative to a set of existing markers.

this we developed a hybrid vision and inertial tracking technique.

The goal of our study was twofold; first we wanted to establish how well users can place annotations in video-see-through AR without any other registered virtual geometry visible. The difficulty with this technique is that no matter how easily users can place the marker at what they think is the correct distance; it is difficult for them to judge how far away the real object is in the first place. The second part of the study addresses this problem by already having some nearby objects labeled for the user. This allows them to judge the new annotation distance *relative* to those already marked objects in the model - a considerably easier task [5].

We first discuss related work pertaining to interaction at a distance in AR and VR, as well as human visual cues for perceiving distance, and other studies that have explored human perception of distance. Next, we introduce the different AR depth cues we implemented for our study. We discuss our system setup, especially with regard improving tracking so that an outdoor study was feasible, followed by a description of our study setup. Finally we will present and discuss the results of our study, and explore possibilities for future work.

2. Related Work

Interaction at a distance has been the subject of much research in both augmented and virtual reality. Much of the work done in VR is not directly applicable to our approach because it assumes complete knowledge of the environment's geometry. In particular, image-plane VR selection techniques, as surveyed in detail by [18], [3], and [4], are not directly applicable to our work. While the same is true for Stoakley et al.'s world in miniature [22], there is some relevance to our work in that we choose as one of our depth cues a top-down view of the scene that represents the 3D cursor's position and any annotations already placed.

We have chosen a different approach than previous work in mobile AR scene model acquisition which also considers far-field outdoor objects [2][16]. While these approaches employ techniques based on different viewing points, we want our users to determine depth from a single point. Similar techniques employed by Baillot et al. [2] and Piekarski et al. [15][16][17] involve "sighting down" feature points or walls of buildings from multiple viewpoints and then using intersections of lines or planes from these views to construct a building. One disadvantage is that this requires walking around the object in question. We are interested in techniques that allow the user to annotate any features that are currently visible, without needing to move. We hope to alleviate the limitation of a static viewpoint by giving the user additional depth information, allowing them to still make accurate distance measurements quickly and easily.

Annotating distant features correctly from a stationary point requires a user to make an accurate depth estimate based on a limited number of natural cues. Cutting [7] proposes that there are five pictorial cues for depth (cf. Section 3), compared to nine for depth perception with motion [6]. Livingston et al. [13] have investigated how some of these pictorial cues, relative size, and occlusion in particular, can be exploited in AR to give the user information about occluded infrastructure. In other work that relates depth perception to augmented reality, Drascic and Milgram [8] focus primarily on how stereo displays contribute to depth perception, while Piekarski and Thomas [17] introduce AR working planes for construction at a distance as a tool to restrict the user's degrees of freedom, allowing them to more easily annotate features on the same plane.

We are particularly interested in being able to move a 3D cursor to a desired location without imposing external constraints (such as a second viewing direction). Previous work has been conducted in this area in VR [11][25][23][10]. Our shadow plane cue is loosely inspired by [10]. Hubona et al. [11] also tested shadows as a cue for object location, albeit in a different manner, casting shadows at random orientations to simulate real shadows, instead of as perpendicular orthogonal projections to be used exclusively as a depth cue.

To increase the accuracy of tracking for our outdoor study we use a hybrid inertial-vision 3-DOF tracking system, inspired by [24][20][12]. Our tracking is most similar to that of You et al. [24], who use point tracking with an inertial tracker to reconstruct 3-DOF pose. The major difference lies in the technique used to combine the inertial and vision components.

3. Depth Cues

The five pictorial depth cues proposed by Cutting [7] include occlusion, relative size, relative density, height in the visual field, and aerial perspective. Of these cues, the first three maintain the same level of quality regardless of distance. Between two objects, it can be determined which one is closer when the distance between them is 0.01% of the distance to the objects using occlusion, 3% using relative size, and 10% using relative density. Height in the visual field is also a useful cue at the depths we conducted our study at: between 20 and 80 meters, but becomes less useful at further distances. Aerial perspective is not of benefit for our intended applications, being most useful for objects on the order of 1000 meters in distance. These numbers, taken from [6], are lower in our case, however, because we conducted our study using video see-through AR. This affects the perception of the real world in two ways. First it limits the resolution that users can see, making relative size and height in the visual field more difficult to judge accurately. Second, it is more difficult to see the whole scene clearly in the fixed dynamic range of a video feed, rather than the high dynamic range of an outdoor environment.

We have developed three different augmented reality cues (see Figure 1) to enhance these pictorial depth cues by giving the user a more clear understanding of objects' spatial relationships, as well as 3D cursor position. The first of our cues projects shadows of the cursor and any placed annotation onto screen-stabilized perpendicular planes at the top and left side of the user's view. Our second cue is a top-down view of the virtual portion of the scene showing any existing annotations as well as the user's cursor. The last cue uses a color encoding and attached dynamic line bars as distance indicators. We change the color of existing marks in the scene based on the cursors depth proximity to each mark. These three cues use progressively less screen space, with the shadow planes taking up over half of the screen (semi-transparently) and the color encoded markers taking up nearly no extra screen space (only the pixels used by the line bars). All of the cues enhance Cutting's five pictorial depth cues. For instance, obstacles do not need to be lined up to check for occlusion because it is immediately apparent in each cue which object is further away. Both, the shadow planes and the top-down view, include tick marks for absolute distance readings. The color-encoded markers extend Cutting's pictorial cues by emulating the effects of occlusion when markers are nowhere near each other, and by simulating an abstract localized version of aerial perspective by changing color based on the distance from the user. We will now explain our AR cues in more detail.

3.1. Shadow Planes

We constructed two head-stabilized semi-transparent planes to cast "shadows" onto from any virtual objects (previous annotations and the cursor) in the scene. Our shadows consist of full color perpendicular orthographic projections of the virtual objects, allowing the shadow walls to be quicker to understand, and perhaps more accurate than standard shadows [11] because the user does not have to decide where the light rays are coming from to determine how the location of the shadow relates to the location of the object [10]. We used two shadow planes because it is sometimes easier to distinguish which shadow belongs to which object on one rather than the other. In our study for instance, since all objects were placed at eye level to eliminate relative height as a depth cue, the top shadow plane was sometimes clearer because on the side shadow plane all objects appear along the same line. To give the shadow planes a sense of absolute scale, as well as showing the relative scale to other nearby objects, we gave the planes a checkerboard pattern, where each check was 5 meters across, as well as having the planes alternate color, and placing distance labels every 10m away from the user.

We also decided made our shadow planes dynamic to use the most useful (high-resolution) part of the plane all the time. Especially when dealing with objects spread over a large distance some parts of the shadow planes, which are parallel to the user's viewing direction, are more useful than others. For instance, if the walls are too far away, then the shadow of a nearby object falls outside of the viewing frustum. However if the shadow walls are too close, then the shadows of two distant objects that are near each other will have indistinguishable depth. At the edge of the user's field of view the greatest disparity from the smallest amount of object depth difference is visible. Therefore we move the shadow planes inward and outward to keep the user's cursor in that useful part of the shadow wall at all times. Figure 1a) shows a typical use of the shadow planes in our study, indicating the light blue cursor, and three out of four dark grey scene markers currently being reflected on the shadow planes. The marker closest to the user has moved out of the range of our dynamic shadow planes.

3.2. Top-Down View

Our second AR depth cue is a top-down view located in the lower left ninth of the screen. This view shows the location of the cursor, as well as any other markers already placed in the scene, and their location relative to the user. This view is also dynamic, changing zoom-level to fully use the space while showing the cursor and all markers in the scene. If no other markers are present, the distance is shown along with the direction to the cursor. We assume that all objects a user may want to annotate from a static point of view are closer than 300 meters. We also show the user a scale with increments every 10m.

3.3. Color Encoded Markers

Our last cue is a simple color encoding of the markers already in the scene to give the user a better relative idea of cursor depth. In this case, the color of each marker changes depending on its depth compared to the current cursor position. If the marker is further away, it will be a shade of green (light green if the cursor is further from it and saturated green as it gets closer). If the marker and cursor are at the same depth the marker will turn red. If the cursor is further away the marker will turn blue, initially saturated blue, becoming lighter as the cursor moves further past the maker. This color scheme gives the most information, through sharp discreet color changes, when the cursor is closest to the marker, the time when that marker is the most important. In addition, each marker has a small attached bar to show the relative distance of the marker to the cursor. A longer bar simply means that the cursor is further away. We include this depth cue because it has the added benefit of requiring little additional screen space. If it performs well it might be a preferable technique, especially for applications where screen space is at a premium.

4. Tracking

To conduct the study accurately, we used a hybrid tracking technique, similar to those presented in [20] [24], combining inertial tracking results from an InterSense InertiaCube2 inertial tracker with vision-based tracking. We used an OpenCV [14] implementation of an iterative Lucas-Kanade method of calculating optical flow for sparse texture tracking. We combined these texture tracked feature points, each of which was originally placed at the location of a known point in the scene with our inertial tracking pose by using an energy minimization function on the three Euler angles returned by the inertial tracker to find the best match with the set of vision points. Because individual vision features can get lost, we paired each feature with a confidence in its location vs. the location of the matching inertial point, updated on a frame to frame basis. If the confidence gets too low, we simply reintroduce the point at its paired inertial marker's location. This reintroduction of points greatly increases the robustness of the algorithm, overcoming many inevitable errors in the vision tracking of individual texture points. In the study, we used a Point Grey Firefly camera with a wide angle lens in full color VGA mode. We used the camera calibration routines from OpenCV to get the correct calibration matrix for our camera, and un-distort each frame before doing vision tracking. We aligned our camera and inertial tracker along the same axis, so that a transform was not needed to go from one coordinate system to the other.

5. Study Layout

We conducted a study to determine how well the above depth cues performed in both the absolute case, i.e. when there are no other annotations in the user's field of view, and relative case. This study took place outdoors, with the users annotating trees, street lights and sign posts that were between 20 and 80 meters away, by placing a spherical marker at the correct location. Objects more than 120 meters away were difficult to perceive given the limited resolution and dynamic range of our camera. We had 19 users participate in our study, 14 men and 5 women between the ages of 23 and 55. Half of our users had never used a head-mounted display before. All users volunteered without compensation. As mentioned before, we used a Point Grey Dragonfly firewire camera, and an

SVGA Sony Glasstron PLM-S700E head mounted display pictured in Figure 2. We used an InterSense InertiaCube2 inertial tracker for the inertial component of our tracking system. For the purpose of this study, no position tracking was used, since users stay relatively static during the trials. Users navigated their cursor through space using an ErgoTouch RocketMouse. Left/right motion was done simply by moving the track-ball left and right, and forward/backward motion was rate-controlled by moving the trackball up and down while holding down the trigger. Users were not able to change the height of the cursor in the study, and all existing markers were also displayed at that same height, eye level.

Before starting our study, we first asked users if they felt they were good judges of distance, and then asked them to estimate the distance to three different objects, without imposing a distance unit on them. One of the three objects was nearer to the user than any objects users were asked to annotate in the study, one right in the middle, and one about 50 meters beyond any they had to annotate in the study. This was done to get an initial idea of how well they judged distance, and to find out what units they thought of distance in. Users were then presented with distances in that unit throughout the study. Roughly an even third of our user population chose yards, feet, and meters respectively as their preferred distance unit in this way.

The first part of the study was to determine how well users could determine distance without any other markers in their view for feedback. Users had to label three different objects approximately 30, 45, and 60 meters, away using four different distance cues. These twelve tests were done in random order, and users were not given feedback on how close their estimates were during the study. The four cues, which users had trained on by labeling a separate medium-distance object with feedback before the study, were the shadow planes and top-down cues described above, as well as a trial where they were given no extra cues, only their 3D cursor, which was a sphere 1 meter in diameter. The last cue was a "depth-aware" cursor. In this cue the cursor would change color, as described in the color-encoded marker cue described above, so that when it was at the correct distance it would turn red. This cue, which assumes that the distance to the object to be annotated is already represented in a scene model, was tested only as a comparison test to see how fast it was possible to make a correct annotation.

The second part of the study consisted of testing cues with other spherical markers present at the correct distance, so relative distance could be used. Again, four cues were tested, and users were first trained with all four cues on a medium-distance object before being tested. Three different test setups were used for a total of another twelve tests. Each setup had four markers already placed

Figure 2: A user participating in the study

in the scene, and users were asked to place one more annotation. The three objects that users had to label were located 38, 55, and 65 meters away. Objects used in the training sessions were never annotated again, so that users could not apply previous knowledge to later tests. The four cues that were used were the three described in Section 3, and another size-only cue. For this, users again had to rely on the size of the cursor, and the size of the other markers in the scene that were also spheres with one-meter diameters.

6. Hypotheses

We formed the following hypotheses before conducting our study:

1. Users would perform only moderately well when given no other visible markers, but the best technique for these tasks in terms of placement accuracy, time, and user preference, would be the top-down view with the size-only cue performing worst.
2. The relative markers would have a strong positive impact on the accuracy of new marker placement versus absolute marker placement.
3. The different cues would have a greater variance in result in the relative case, with the shadow planes and top-down view performing the best, in terms of placement accuracy, time, and user preference

7. Results

Our results differed from what we expected. A number of our hypotheses agreed with our empirical findings, but there were also a number of important differences. We will first discuss the absolute case results, followed by the relative case results.

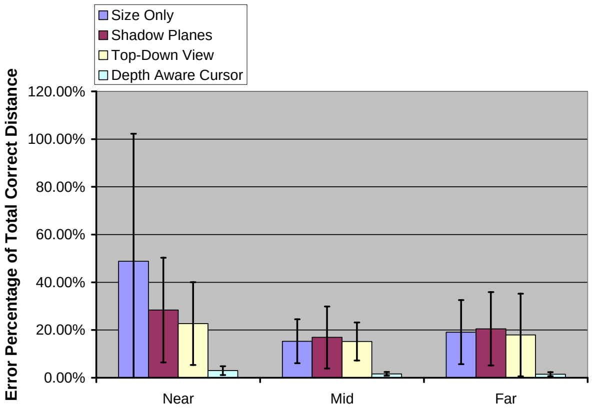


Figure 3: Results of absolute tests with different object location and no other markers present. Size-only performs worse in the near condition.

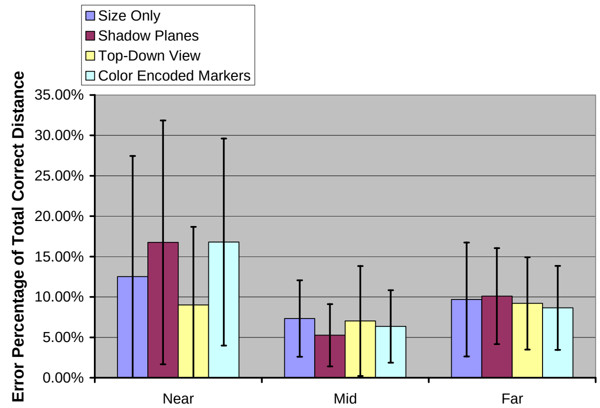


Figure 4: Results of relative tests with four different cues, again split up by average object location.

7.1. Absolute Case

The different cues in the absolute case did not quite perform as we had anticipated; There was no statistically significant difference between the three techniques (ANOVA, p=0.23). In retrospect, this is not entirely surprising since in absence of relative depth cues, we mainly tested how well people were able to judge distances in video see-through AR. For example, while the user could tell more easily that the cursor was 50 meters away in the top-down view than in the size-only cue, that may not have been the limiting factor. Instead, the limiting factor was probably the users' estimation of how far away the physical objects were, and in this they only had experience, and the size of the cursor (which was present in all cues) to help them.

There was an improvement for the two cues that did offer more information (top-down view, shadow walls), albeit not a statistically significant one. Users had 27.7% error (meaning if the correct location was at 100 meters users were off by 27.7 meters on average) for the cursor size-only cue, and 21.9% and 18.6% for the shadow planes, and top-down view respectively. However, because of the large amount of variance between users, and even for a single user from one test setup to the next, this was not a statistically significant difference. Figure 3 shows the results for all cues, grouped by the three different distances we tested them on. It is important to note that the error displayed is relative to correct distance, so although the percentage of error is much higher for objects that were nearby, the absolute amount of error was similar to that in the far case. These error ratios were slightly less than in the initial questioning phase when the users were not wearing a HMD, likely because users were told the range of the objects they were marking after they completed these initial questions.

The times that the different techniques took were in the range we expected them to be, but again, much closer together than we expected. The distance-aware color changing cursor cue took 15.9 seconds for users to move to the correct location, while the other three cues made the annotation task take 26.5, 32.4, and 30.2 seconds for the color only, shadow planes, and top-down view respectively. All of these times also had a large amount of variance (standard deviation of approx. 15 seconds for each one), so there was no statistical significance between these times. There was a statistically significant effect of cue repetition on time (ANOVA, p=0.0029), but not on accuracy. This means that there was no learning effect (since no feedback was given during testing), but users grew more confident in their use of the cues.

There was a statistically significant difference in how users perceived how easy it was to move the cursor to the correct distance using the different techniques (ANOVA p=0.0011) as can be seen in Figure 6. The top-down view was the clear favorite, being preferred to both the shadow planes (PAIRED T-TEST p=0.035), and the cursor size-only cue (PAIRED T-TEST p<0.0001). The color encodings received similar high satisfaction ratings.

7.2. Relative Case

Having the other markers visible did overwhelmingly improve (ANOVA p<0.0001) the results, reducing error from an average of 22.74% for all of the absolute cues combined to 9.89% for all relative cues combined (see Figure 5 for a graph that splits it up by cue). Against our expectations, no technique established itself as a clear winner. However, there was one case, which can be seen in figure 5, where we found a significant difference between any two techniques. In the near trial the top down view gave a statistically better result, 9.02% error, vs. 16.80% error (PAIRED T-TEST p=0.016) than the color coded marker cue. The size-only technique had 9.85% error through the three trials, while the other

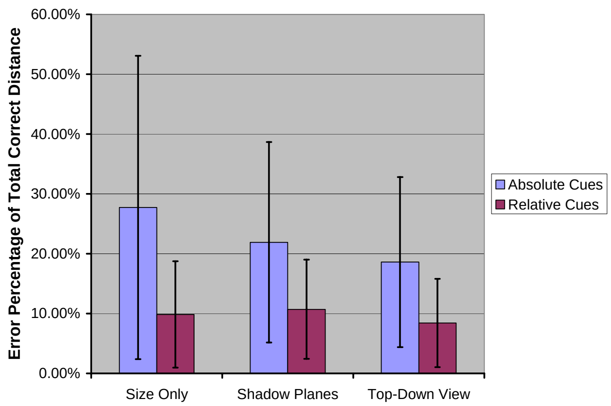


Figure 5: Results with and without other markers present. Relative annotations had significantly less error.

techniques had 10.60%, 10.71%, and 8.42% error for the color-encoded markers, shadow planes, and top-down view cues respectively. We expected the shadow planes, and top-down view to do significantly better, especially than the size-only cue. There were a couple of contributing factors to why we think the techniques performed rather similarly.

First, users were limited by how well they could see the real world through the video see-through display. Thus, even though they could see where other trees from the virtual model were in the real world, they were unable to tell in the real world the proximity of those trees. This was mentioned as a problem by a number of users, especially when trees were far away. One user exclaimed upon taking off the HMD after the study "Oh that's where those trees were." The quality of the camera was definitely a contributing factor to this, being designed more for size and weight than video quality.

The second possible factor is from the way people decided on the distance to objects. From observing people in the trial it seemed that most people would find one nearby object that already had a virtual marker, and then make the binary decision if the object they were supposed to mark was closer or further than that. Then from there they would place their new annotation slightly closer or further away. For all of the techniques it is easy to tell when the cursor is at the same depth as another virtual marker (from occlusion if nothing else), and from there moving the cursor slightly closer or further might have been treated by the users as independent from the depth cue. Users might have resorted to this technique because of the lack of clarity in the video stream. It was much easier to tell which object was closer, than how much closer it was.

Interestingly there was again a statistically significant difference in how the users felt about how easy it was to judge distance using the different techniques (ANOVA $p=0.0011$), with the size-only cue again being the least

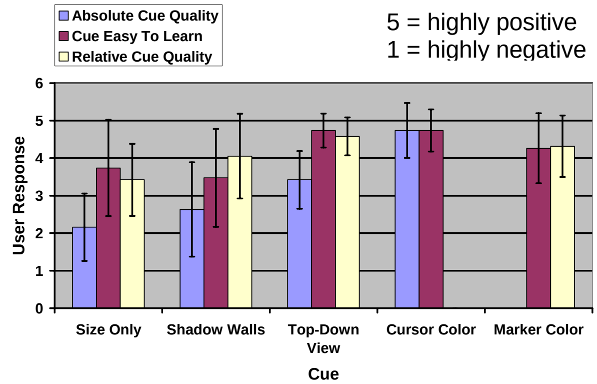


Figure 6: User Questionnaire Results. Users preferred Top-down View.

favored, and the top-down view being the favorite. People also liked the color encoded markers better than the shadow planes by a small amount, possibly because some people had difficulty seeing the shadows clearly in the bright outdoor environment. Times for the different techniques were very similar to those from the absolute case, averaging about 27 seconds per trial. The slowest was the shadow planes at 29.2 seconds, and the fastest was the size-only cue at 24.6 seconds. There was no statistical significance between them, but there again was an inverse effect of repetition on time.

8. Conclusions and Future Work

While our results were not in total agreement with our hypotheses, we still arrive at a number of interesting findings. First and foremost, relative cues were a great aid in determining depth, cutting error values approximately in half. The top-down view came out as the preferred technique to provide distance information to the user, both in the questionnaire and in accuracy, albeit only statistically significant in the trials where the target was nearest. Surprisingly, shadow planes and the top-down view were not a highly significant help to users in refining their distance estimates significantly compared to the size-only cue. We feel that this is because of a combination of factors, including the use of video see-through and its visibility limitations. The screen-estate-saving color-encoding technique performed well overall. With other markers in the scene, users were able to place a new mark with only 10% error, which could be sufficient for many applications.

In further studies, we are very interested in determining application preferences for when to use any of the following three techniques for determining the 3D position of a new physical feature: GPS footprint, intersection of planes [17], and distance estimation from a static point of view as explored here. We hope to also decrease our error rate further through various ideas. One simple idea would be to give a choice of different objects for the cursor. Because the cursor had a physical size, it bridged the gap between the physical and virtual scenes. We used a one-meter diameter sphere as our cursor, but would like to find out if users could more accurately tell how far away it was if it was a more recognizable object of a certain size. For instance if it was shaped like a bicycle, or a car, would users have an easier time telling what that recognizable object would look like at different distances? We would also like to try to improve the estimations given by the user automatically by extending our existing vision-based tracking. If multiple features are being tracked, parallax from user motion could be used to tell the relative depth to those objects, to refine their positions. Since the user's position can be found with GPS, refining the object placement would depend on the distance walked and position-tracking accuracy. This would be an interesting hybrid technique between static viewpoint estimation and ray intersection from radically different viewpoints

9. References

[1] Abowd, G.,C Atkenson, J. Hong, S. Long, R. Kooper, and M. Pinkerton, Cyberguide: A mobile context-aware tour guide. In *AMC Wireless Networks*, 1997, pp. 421-433.

[2] Baillot, Y., D. Brown and S. Julier, Authoring of Physical Models Using Mobile Computers. In *ISWC '01*, Zürich, Switzerland, 2001, pp. 39–46.

[3] Bowman, D. A. and L. F. Hodges, An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *AMC I3D '97*, New York, 1997, pp. 35-38.

[4] Bowman, D.A., Interaction Techniques for Common Tasks in Immersive Virtual Environments. *PhD thesis*, Georgia Institute of Technology, Atlanta, GA, 1999

[5] Chapanis, A. and R. McCleary, Interposition as a cue for the perception of relative distance. *American Journal of Psychology*, (48), 1955, pp. 113–132.

[6] Cutting, J. E. and P. M. Vishton, Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In *Handbook of Perception and Cognition, Vol 5; Perception of Space and Motion*, Academic Press, San Diego, CA, 1995, pp. 69–117.

[7] Cutting, J. E. Reconceiving Perceptual Space. In *Perceiving Pictures: An Interdisciplinary Approach to Pictorial Space*, Cambridge, Ma, MIT Press, 2002.

[8] Drascic, D. and P. Milgram, Perceptual Issues in Augmented Reality. In *SPIE Volume 2653: Stereoscopic Displays and Virtual Reality Systems III*, San Jose, Ca, Jan 1996, pp 123-134.

[9] Feiner, S., B. MacIntyre, T. Höllerer, and A. Webster, 1997: A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. In *ISWC '97*, Cambridge, MA, 74–81.

[10] Herndon, K. P., R. C. Zeleznik, D. C. Robbins, D. B. Conner, S. S. Snibbe, and A. van Dam, 1992: Interactive shadows. In *ACM UIST* '92, 1–6.

[11] Hubona, G., P. Wheeler, G. Shirah and M. Brandt, The Relative Contributions of Stereo, Lighting, and Background Scenes in Promoting 3D Depth Visualization. In *AMC Transactions on Computer-Human Interaction*, Vol. 6 Issue 3, ACM Press, New York, 1999, pp. 214-242.

[12] Jiang, B., S. You, U. Neumann, A Robust Tracking System for Outdoor Augmented Reality, In *IEEE Virtual Reality*, Chicago, March, 2004, pp. 3-10

[13] Livingston, M. A., J. E. Swan II, J. L. Gabbard, T. Höllerer, D. Hix, S. J. Julier, Y. Baillot, and D. Brown, Resolving multiple occluded layers in augmented reality. In *ISMAR '03*, Tokyo, Japan, 2003, pp. 56-65.

[14] OpenCV, www.intel.com/research/mrl/research/opencv

[15] Piekarski, W. and B. H. Thomas, Tinmith-Metro: New Outdoor Techniques for Creating City Models with an Augmented Reality Wearable Computer. In *ISWC '01*, Zurich, Switzerland, Oct 2001, pp 31-38.

[16] Piekarski, W. and B. H. Thomas, Interactive Augmented Reality Techniques for Construction at a Distance of 3D Geometry. In *7th Int'l Workshop on Immersive Projection Technology / 9th Eurographics Workshop on Virtual Environments*, Zurich, Switzerland, May 2003.

[17] Piekarski, W. and B. H. Thomas, Augmented Reality Working Planes: A Foundation for Action and Construction at a Distance. In *3rd Int'l Symposium on Mixed and Augmented Reality*, Arlington, Va, Oct 2004.

[18] Pierce, J. S., A. S. Forsberg, M. J. Conway, S. Hong, R. C. Zeleznik, and M. R. Mine, Image plane interaction techniques in 3d immersive environments. In *ACM 1997 Symp. on Interactive 3D Graphics*, 1997, pp. 39–43.

[19] Rekimoto, J., Ayatsuka, Y., and Hayashi, K. (1998). Augment-able reality: Situated communication through physical and digital spaces. In *ISWC '98*, pages 68–75

[20] Satoh, K., M. Anabuki, H. Yamamoto, H. Tamura, A Hybrid Registration Method for Outdoor Augmented Reality. In *2nd Int'l Symposium on Augmented Reality*, New York, Oct. 2001, pp. 67-76

[21] Spohrer, J. WorldBoard, www.worldboard.org/ pub/spohrer/wbconcept/default.html, 1997

[22] Stoakley, R., M. Conway, and R. Pausch, Virtual reality on a WIM: Interactive worlds in miniature. In *ACM CHI '95*, 1995, pp. 265–272

[23] Wither, J., and T. Höllerer, Evaluating Techniques for Interaction at a Distance. In *Int'l Symposium on Wearable Computing*, Arlington, VA, 2003, pp. 124-127.

[24] You, S., U. Neumann, and R. Azuma, Orientation Tracking for Outdoor Augmented Reality Registration. In *IEEE CG&A*, Vol. 19, No. 6, Nov. 1999.

[25] Zhai, S., W. Buxton, P. Milgram, The Partial-Occlusion Effect: Utilizing Semitransparency in 3D Human-Computer Interaction. In *ACM TOCHI*, Vol. 3, Issue 3, New York, Sept. 1996, pp. 254-284